Overconfidence via Self Assessment:

Prediction Accuracy & Pre-test Self-Assessment Survey Administration

John W. Lee

Southern New Hampshire University

Author Note:

John W. Lee is a graduate student in the Master of Science in I/O Psychology program.

This Capstone Project represents a research proposal presenting an introduction to the chosen research topic, a literature review, research question, methods, and conclusion.

Introduction/Problem

Orators from Benjamin Franklin to the Delphic Oracle have pondered the importance of self knowledge (Dunning et al., 2018). Many of those speculations have concluded, along with a wealth of psychological research, that accurate knowledge of self is tantamount with success (Dunning et al., 2004; 2018). Within the process of self-evaluation, and self-assessment, consistent integral flaws have been observed both in the laboratory as well as in the field (Dunning et al., 2004). Overconfidence considered as a phenomenon visible in flawed self-assessments has been widely studied (Dunning, Heath, & Suls 2004, 2018; Freund & Kasten 2012; Logg, Haran, & Moore 2018, Zell & Krizan 2014). As a method to produce or provide the opportunity for self-evaluation and self-assessment, survey based self-assessments are commonly used to capture ideas, tendencies, traits, and personalities reflective of the self (Dunning et al., 2004; 2018). These surveys often represent an opportunity to examine individual attributes and characteristics reflective of a particular individual, or group of individuals. The extent that inferences are drawn from self-assessments surveys, the manner in which those assessments surveys are conducted, and the assessed validity of the self assessment, can all depend on the ability of the researcher designing the assessment to account for certain phenomena, one of them being overconfidence (Dunning et al. 2018).

The ability to gauge the level of confidence held by a particular respondent is valuable to organizations interested in understanding their employees, leaders, or even clients (Dunning et al., 2004). Additionally, in certain scenarios, overconfidence can be a trait organizations wish to avoid in their leadership, or prospective employees (Dunning et al, 2004). Specifically, overestimation within organization's can cause damage to the

individual and the organization (Dunning et al. 2018), whereas the ability to improve the level of accuracy displayed stands to benefit the organization and the individual. With this in mind, designing a measure that include facets conducive with the identification and potential reduction of phenomena like overconfidence stands to benefit organizations with an interest in the accuracy of it's members, as well as psychological researchers interested in learning more about the cognitive nature of self assessment.

The research question I propose is the following: Will volunteer respondents solicited through an established online survey participation forum accurately predict their score on the BAROCO short syllogism test (Shikishima et al., 2011a), and does a correlation exist between the accuracy of prediction and the aptitude measured. The second part of this research question is that when the pre-test survey questions are administered utilizing a pre-recorded video of the survey administrator asking the pre-test self-assessment survey questions, will respondents be less likely to display overconfidence than if the pre-test self assessment survey were administered via a standard computer generated text survey.

Hypothesis #1. Respondents with elevated scores on the BAROCO short syllogism test will have an accuracy of prediction greater than the accuracy of prediction of respondents with lower scores on the same test.

Hypothesis #2. Respondents given the pre-recorded video version of the pre-test self-assessment survey will display an accuracy of prediction greater than the accuracy of prediction of respondents given the text only version of the same pre-test self assessment survey.

By examining the relationship between the level of confidence expressed in regards to performance on an intelligence test or a measure of general ability like the BAROCO Short, this research accomplishes a two-fold purpose. Firstly, if the prediction that overconfidence and the score on the BAROCO short are correlated, and that correlation is positive, this research affirms prior research into overconfidence, and over placement regarding specific characteristics of overconfidence. Here, that affirmation would be that the better one does on an objective measure like an intelligence test, the more accurate they are at predicting how well they will do. This information is valuable to psychologist who are conducting research that utilizes self assessments and has an emphasis on self awareness, or who's research could be impacted by the phenomenon of overconfidence.

The second purpose achieved through this research is the creation of a reliable measure to establish a baseline level of overconfidence through the use of a simple, easily administered and accessible test that when coupled with a brief pre test assessment survey, provides an accuracy of prediction score for each respondent. This information has the potential to be valuable not only to future psychological research, but also has the potential to be immediately applied within organizations who wish to gauge a perspective employees, perspective leader, or even a client's accuracy of prediction. Because our wealth of knowledge and available research into overconfidence is large, this information can be used in a number of ways to formulate other predictions regarding individual traits, or abilities.

In addition to the gains achievable in regards to the simple identification of overconfidence, this research purposes novel and textured ideas that synthesize years of

research into the phenomenon of overconfidence in order to develop a prediction that seeks to unravel what has been identified by some researchers as a gap in our understanding of the motivations behind overconfidence (Dunning, et al. 2004). Although these elements of this research are limited to the motivational theories already established in regards to phenomenon such as overconfidence and motivation, the prediction that survey administration method can have the ability to alter motivation has the potential to offer new, and unchartered insights into millions of pages of survey data collected since the introduction of computers into the world of psychological science, and the administration of self assessments. Here, the prediction that by introducing a face to face condition during a self assessment, motivation can be moderated as reflected in a significant change to a phenomenon such as overconfidence, opens the door to the role of computer based survey's in self assessment data collection, and the identification of phenomenon such as overconfidence.

One of the limitations of the proposed research is that in exploring the correlation between accuracy of prediction and general ability, the measure of general ability may not contain enough data to formulate a reliable indicator. One way to address this, is through the use of the demographic question regarding level of education, where level of education has been found to correlate with level of general ability as demonstrated on the BAROCO short (Shikishima et al., 2011a). An additional limitation of the proposed research method is by utilizing MTurk, the level of balanced groups obtained that are representative of the general population in each of the geographic areas could be limited based on the current active users, and those users inclusion in representative groups. Every effort will be made to achieve a wide sample of respondents within the survey

respondent forum, that is representative of the general population of people between the ages of 18 and 68.

The findings of this study will be solicit for interest from publishers of psychological journals who's topics include self assessment, organizational development, Industrial/Organizational Psychology, and others. In addition to publication in scienctific journals, the results here will be made available to those within the industry of organizational consulting, with a possible end goal of an easily administered level of overconfidence measure being made available for commercial use. The insights gained regarding the impacts of overconfidence can be disseminated to applicants within the workforce, students, managers, and researchers in a wide variety of industries, subjects, and markets. This can be accomplished through the use of the associated presentation to gain interest, solicit for research partners, or offer ideas on the identification of overconfidence.

It is worth mentioning here that after conducting an extensive review of the available psychological literature regarding overconfidence, all most all instances of overconfidence identified in experimental settings were obtained either through the use of pen and paper administered surveys, or computer based surveys, with almost no identification of the phenomenon of overconfidence found through the use of in person interviews or similar. Although not specifically addressed within this research, where limitations exists regarding the need for observable, and theoretically supported research frameworks, further research into the role of in person interviews in the identification of phenomenon such as overconfidence in self assessments is believed to be justified.

In order to fully understand the problem of overconfidence as it is displayed in self assessments, and how that overconfidence can be damaging to an organization, it is valuable to consider self-assessments and the impact overconfidence can have, an operational definition of overconfidence, self-assessment administration methods, theory associated with the motivation behind overconfidence, and a particular opportunity for overconfidence to be displayed.

Literature Review

Self Assessments

David Dunning from the University of Michigan, Chip Heath from Stanford University and Jerry Suls from the National Cancer Institute (2018) published an in-depth review of self-assessments, and specifically the impacts of erroneous self-judgments on health, in the workplace, and within the field of education. The author's message was that an individual's skill and their actual ability correlate only moderately at best, and sometimes not at all and that people often over place themselves in relation to other people (Dunning et al. 2018). The authors point to the 2014 work by Dunning & Helzer that found that those who show poor performance have marked overconfidence, and those who perform well have a much better accuracy in judging themselves (Dunning et al. 2018).

Dunning et al. (2004), offers a unique monograph on self-assessment that explores years of research identifying empirical evidence highlighting flaws in self-assessment. These flaws are described as pervasive and widespread throughout the research, and impactful on outcomes in multiple domains including organizational performance (Dunning et al., 2004). Here, a flawed self-assessment is described as the

human being's attempt to describe, project, or otherwise display an assessment of themselves, or of their future performance, in a way that is inaccurate (Dunning et al., 2004). This allows for the introduction of overconfidence as displayed during a self-assessment to be described as an example of a flawed self-assessment.

The regularity with which Dunning et al., (2004) is able to identify study after study, including laboratory based research and real world scenarios, in which people consistently and erroneously predict their future performance, is remarkable, as well as indicative of the extensive literature available regarding the phenomena of overconfidence (Dunning, Heath, & Suls 2004, 2018; Freund & Kasten 2012; Logg, Haran, & Moore 2018, Zell & Krizan 2014). This is not surprising, given that self appraisals and knowledge of self have been explored both in psychological research, as well as within literature and historical accounts that in some cases date back centuries such as the inscription on Apollo's Temple that read "Know Thyself", or Laozi statement in Sixth Century BCE China who philosophized that "knowing others is intelligence, but knowing yourself is true wisdom" (Dunning et al., 2018). In addition to establishing the historical emphasis on self-knowledge, the importance of accurate self-appraisals and assessments has been identified through research that describes negative outcomes for individuals, organizations, or other affected parties, when self-assessments are incorrect (Dunning et al., 2018).

Dunning et al. (2004) presents a systematic and thorough comparative analysis of the laboratory findings as well as real world examples that explore the struggle of the individual to produce an accurate self-appraisal. The authors argue that the capacity held by most individuals to evaluate themselves or predict their own behavior is often times

much more modest that what intuition would commonly have one believe (Dunning et al., 2004). This research provides evidence that supports the observation that accurate self-judgments involve complex forces that influence both self behavior and social behavior (Dunning et al., 2004). This alludes to the difficulty in obtaining accurate self knowledge when it is desired, for example a pre-employment survey relied upon to provide organizations with information regarding the applicant that reflects the actual capabilities of the individual. Here, it is clear to see potential stakeholder benefit to the organization in receipt of valid a reliable data produced from a self-assessment conducted as part of pre-employment screening that produces information regarding actual abilities, rather than conjecture that has a high probability of being incorrect. The complexity involved in making self-judgment's and associated erroneous self-evaluations often produced as identified by Dunning et al., (2004), could help to explain the counter intuitive findings that display repetitive flaws in self assessments, specifically regarding people's overconfidence.

The monograph by Dunning et al., (2004) points to empirical evidence regarding flawed self assessment specifically relating to self-assessments of knowledge, or intelligence. By evaluating prior meta analyses conducted within the field of self assessments, the authors support the position with a wealth of evidence that often a weak correlation, or even a null correlation exists between objective performance and knowledge or skill (Dunning et al., 2004). The psychological mechanisms underlying flawed self-assessments described by Dunning et al., (2004) are given in relationship to two very common biases that have been widely studied, those of above average effects as well as overestimation of the likelihood of desirable events (Dunning et al., 2004). The

Authors explicitly state that the forces that provide influence to self-behavior and social behavior are complex, and that self-assessments have been shown to be systemically flawed, imperfect, and that those flaws carry significant implications within numerous organizational circumstances (Dunning et al., 2004).

Overconfidence and Over Placement

Building on the research conducted by Dunning et al., (2004), Jennifer M. Logg from Harvard University, Uriel Haran from Ben-Gurion University of the Negev, and Don A. Moore from the University of California, Berkeley (2018) provide an extremely relevant definition of flawed self-assessment in stating that these flaws can be described as over placement. Over placement exists when a person exaggerates their own position as it relates to others, predicting that they will perform better on a test than they actually perform (Logg, Haran, & Moore, 2018). The researchers provide an experimental design aimed at understanding the causes for overconfidence by comparing the effect of both motivation and cognitive errors on over placement (Logg, Haran, & Moore, 2018). In a series of experiments, the authors explore estimating personal intelligence (Logg, Haran, & Moore, 2018). The self-assessment involved in predicting personal outcomes on intelligence is described as exploring over placement when the outcome is objective and verifiable (Logg, Haran, & Moore, 2018). This establishes intelligence testing as an objective and verifiable assessment that is likely to elicit over placement (Logg, Haran, & Moore, 2018).

The methods chosen to explore this particular facet of over placement and overconfidence in self assessments saw a group of 111 men and women made up of students and faculty or staff at a University in the East. This study included a between

subjects design where participants were asked to estimate their performance both before and after the short test of math and logic, and were also placed into a motivation condition that was manipulated to evaluate motivation's impact on both outcomes, as well as over placement (Logg, Haran, & Moore, 2018). The findings of this particular experiment provided results consistent with the large majority of empirical support that predicts that overconfidence can be minimized by regulating the difficulty of the tasks, with tasks that are more difficult eliciting less overconfidence, and tasks that are less difficult eliciting more overconfidence (Logg, Haran, & Moore, 2018).

The results of this set of experiments presents novel findings and makes the assertion that motivation impacts overconfidence less than prior research has established (Logg, Haran, & Moore, 2018). These findings point toward an explanation that the overconfidence displayed regarding performance on an upcoming IQ test, did not arise from a motivation to view ones self as intelligent, and instead arose from some other motivational influence (Logg, Haran, & Moore, 2018). When considering the complexity of forces acting on self assessments discussed by Dunning et al. (2004), this new insight into the consistent flaws expressed during self assessments, opens the door to investigate other forces that may be producing overconfidence in self assessments (Logg, Haran, & Moore, 2018).

Motivation Behind Overconfidence

In regards to motivation and overconfidence, Self Determination Theory (SDT) presents a helpful tool for contemplating patterns like overconfidence that are found among self assessments, such as with psychological assessments administered following treatment (Klag et al. 2010). SDT posits that some behaviors are more likely than others

to appear when intrinsic, as opposed to extrinsic motivators, are utilized (Ryan & Deci, 2000). Where Logg, Haran, & Moore, (2018) established that overconfidence may not be an indicator of a persons motivation to maintain a positive self view, introducing SDT establishes a set of values that can be observed in relationship to motivation that helps to focus the lens of the complex forces in play during self appraisal (Dunning et al., 2004), and specifically self appraisal of the outcomes on an intelligence test (Ryan & Deci, 2000). With this in mind it is plausible that the motivation behind overconfidence displayed during self-assessments could then be impacted if specific sources of motivation are altered or changed through the administration of the survey itself. When the particular method of survey administration is changed toward extrinsic motivation, extrinsically motivating factors may then be more likely to be elicited. Similarly if intrinsic motivation is incorporated into survey administration, it seems plausible that intrinsically motivated factors would become more prevalent. This application would then rest on the assertion that an anonymous computer generated survey is described as engaging intrinsic motivation, where the only known incentive to provide an accurate appraisal during the short time a pretest self-assessment survey is administered, is intrinsic motivation (Ryan & Deci, 2000) This is in contradiction to a face to face pretest survey administration, introducing another person into the self assessment process and offering the opportunity for extrinsic motivation to impact the likelihood of certain behaviors being displayed, like overconfidence (Ryan & Deci, 2000) Here, we introduce examples of potential extrinsic motivation such as the desire to please, or otherwise controlled motivation such as accountability (Ryan & Deci, 2000). SDT offers the opportunity to specifically describe the motivation behind the predicted outcome of a

pre-test survey administered in person, as opposed to one administered via text on computer screen.

Self Determination Theory holds that even though intrinsic motivation is extremely important, it is far from the only type of motivation (Ryan & Deci, 2000). Self-assessment presents the opportunity for separate, distinct and complex motivational forces to be introduced (Ryan & Deci, 2000; Dunning et al., 2004). If it is possible to determine that there are extrinsic motivators present during self-assessments, and these extrinsic motivators are capable of reducing certain phenomena such as overconfidence, then knowledge on necessary steps to mediate or reduce overconfidence is gained. This knowledge could allow the development of effective warnings that utilize extrinsic motivation in order to effectively reduce overconfidence. Additionally, by demonstrating that the existing body of knowledge regarding overconfidence contains a reliance on intrinsic motivation as the foundational theory, where challenges to current understanding of overconfidence based on similar motivational theories are in question (Dunning et al. 2018), this research stands to open the door to the role of the computer in psychological assessments and what motivations are being triggered through their use.

Self-Assessment Administration Method

Mick Couper and Benjamin Rowe (1996), from the University of Michigan conducted research at the Survey Research Center to investigate reactions and performance from respondents on a self-administered personal interview survey. The researchers here sought to find data regarding the reaction of respondents to the use of a computer to administer a survey (Couper & Rowe, 1996). The author's utilized a self-portraits study to examine differences between those who took the survey in the strict

face to computer screen condition, or the face to face condition, and compared the quality of the data produced by each condition (Couper & Rowe, 1996). The findings established differences regarding the substantive responses that were given in the face to computer condition and a subsequent face to face administered item.

The authors of this study provide methodological frameworks that demonstrate differences in self-assessments administration methods, and the associated outcomes. Although overconfidence is not directly addressed, the specificity to which the authors explore specific aspects of the substantive content differences between survey administration types, speaks directly to a link between a standard text survey administration format, and one that involves more stimulation and engagement, such as the use of a pre-recorded video of a live person as a means to administer the self-assessment pre-test survey. The rapid acceleration over time of the use of computers by a large majority of the world's population culminating in 2020 where computer usage is widespread, the historical validity of the questions posed regarding the role of the computer offers an unadulterated view of that interaction, somewhat removing the computer acculturation that has occurred both within, and outside of the realm of psychological research since the time the study was published.

The BAROCO Short Syllogism Test

In discussing self assessment, overconfidence and it's motivators, it seems important to also consider ways in which overconfidence can be elicited in a self-assessment. Intelligence tests, academic exams, performance on upcoming sports events, have all been used as measures to afford respondents the opportunity to display overconfidence, or over placement (Dunning et al., 2004). Distinctions have been made

regarding the type of measure, test, or upcoming events, and the level of overconfidence commonly displayed (Logg, Haran, & Moore, 2018). Objective measures, such as intelligence tests, have been found to produce measurable levels of overconfidence (Logg, Haran, & Moore, 2018). The BAROCO short is a syllogism test comprised of five syllogism solving questions, with findings substantiating a close relationship between the ability to solve syllogistic problems and general ability (*g*) (Shikishima et al., 2011b). The BAROCO short has been used to identify *g* across multiple populations, ranging in ages, socioeconomic status, gender, race, and has also been studied across cultures (Shikishima et al., 2011b). The test has been established as a valid and reliable measure of *g*, and found to be capable of highlighting differences similar to other measures of intelligence and academic ability (Shikishima et al., , 2011b). Because the test utilizes abstract concepts and vocabulary, it is not easily impacted by attempts to "cheat" by looking up answers making it an ideal candidate for inclusion in psychological research conducted outside of the laboratory regarding self-assessments. In addition to syllogistic reasoning being found to be an accurate predictor of *g*, reasonable support was established for the construct validity of the BAROCO Short as a proxy intelligence test (Shikishima et al., 2011b), making it an excellent choice for an objective and verifiable measure of intelligence in relationship to self-assessments and overconfidence.

Research Design

This study endeavors to analyze and observe in part how accurate people are at predicting their score on a specific test, here the BAROCO short syllogism test is used. Once a level of accuracy is established for all respondents, those respondents can then be compared based on their test scores, in order to determine if there is a correlation between

the level of accuracy of prediction and the actual test score. This part of the research incorporates a sample population of survey respondents, balanced for age, education level, ethnicity and geographic location whom first take a brief pre-test self assessment survey, followed by the BAROCO Short syllogism test in order to observe the presence of overconfidence in their assessments.

A randomized experimental design is proposed that explores whether or not pre-test self-assessment survey administration method has an impact on overconfidence in two randomly assigned groups. The experimental between subjects design utilizing random assignment to one of two conditions is chosen in order to demonstrate future support for the findings related to overconfidence expressed in relationship to an objective measure of intelligence, and attempt to establish greater validity in the results (Trochim, 2020). In order to facilitate these questions, a quantitative data analysis is accomplished through coding. By coding the self assessment responses into quantitative representations, it allows for detailed data analysis, establishment of statistical significance, and comparisons of individual variables (Trochim, 2020).

The strength of this hybrid experimental design is by utilizing a measure of general ability, like the BAROCO short, the opportunity to directly observe overconfidence in self assessment is created. Additionally, by using an objective measure of intelligence like the BAROCO short, past research regarding overconfidence specifically relating to objective measures is built upon, providing further support for findings. The use of the between subjects experimental design where respondents are randomly assigned to one of two versions of the pre-test survey administration method will strengthen the internal validity of the results found.

Participants

The respondents in this study will be solicited through Amazon Mechanical Turk (MTurk), where an ad as requester will be placed for respondents between the ages of 18 and 68 to complete the assessment in exchange for payment. The length of the research survey for the respondent will vary between 3mins to 10mins, making the average time 6.5 minutes. In order to determine fair compensation for respondents, the average hourly wage commonly accepted in the region where the respondent is living will be utilized. In the United States the total private average hourly wage of nonfarm workers, seasonally adjusted, as of June 2019 is $29.37 (BLS, 2020). Dividing $29.37 by 60 minutes equals (29.37/60 =) $.49cents per minute. Because the average time of completion is projected at 6.5 minutes, we then multiple $.49cents per minute by 6.5, (6.5x.49) to get $3.19 for 6.5 minutes of work. This amount will be the compensation for completing the assessment, where each respondent choosing to participate via MTurk, and residing in the United States, will receive $3.19. Respondents solicited from other countries will be given the compensation based on the same calculation, utilizing the closest geographic region that has economic information sufficient to produce a fair wage for the completion of the survey. The survey will be open for a period of 3 months in order to increase the number of respondents. Standard demographic questions will be asked to provide the opportunity for respondents to disclose their age, ethnicity, and highest level of education. Gender was specifically left out of the demographic questions to avoid any perceptions related to socially established beliefs related to gender and intelligence (Musto, 2019). The ideal sample size will be balanced for all demographics and include a large enough number of respondents to provide the opportunity for establishing statistically significant results.

Materials

The assessment will be administered via the computer, requiring that respondents have access to a computer with internet or similar device capable of playing a short 2minute, pre recorded video stream, general knowledge of how to operate a computer, and familiarity with the MTurk software. The assessment is written in English, as well as at least one predominate language from the Middle East, and one from the East. Respondents will be required to select their desired response to questions generated through MTurk by selecting the correct on screen key or button. Timing measures will be included to limit the amount of time a respondent can spend on any given question, as well as to assess the average amount of time taken on each question. This is in order to limit respondents from looking up answers, or information, prior to selecting a response, and will ensure that the one minute recommended time limit per question, established by the BAROCO short syllogism test, is maintained.

The pre-test self-assessment survey will be arranged into two different versions, version one and version two. In version one, the materials will include a text welcome message, text instructions, text demographic questions, text general prediction question, followed by the BAROCO short with individual prediction questions and assessment completion message, all administered via text appearing on the computer screen, with no pictures, or associated animation or graphics (appendix 1). In version two, the assessment will include a pre-recorded video that corresponds to text displayed on the screen (appendix 2). The video will be preceded by text that states: “let us introduce you to __(common male name per region or common female name per region)__, they will be administering your assessment today. We would like to make your interaction with the

test administrator as realistic as possible so please look directly into the eyes of _____ while they are talking, and mark your answer only after they have finished talking." The video will consist of a person seated behind a desk, looking directly into the camera and stating the exact same text, including the welcome message, instructions, demographic questions, general prediction question, and survey completion message as stated in version one. The BAROCO short will still be entirely administered via text appearing on the screen, however each of the individual prediction questions will be administered via the video recording. The video will pause while a respondent is providing an answer. When arriving at the demographic questions, the answer selections will be shown on the screen, and the respondents will be asked to make their selection on the screen following the completion of the person talking.

In order to ensure the internal reliability of this element of the assessment, four total video's per designated language will be created, two from female actors and two from male actors, all within the median age of the range 18 to 68. The two videos from female actors and the two videos from male actors will be given to a group of similar volunteers solicited through MTurk, along with opportunities for respondents to provide answers. This will be conducted prior to commencing the study, and will be done to establish the reliability of the videos, through analyzing the variance produced by each video. A smaller sample size will be utilized here, with the minimum amount of respondent's utilized necessary to obtain valid results. The videos that display the lowest level of variance from respondents will be selected as the final videos to be used in the study, resulting in one male video, and one female video for each language the assessment is conducted in. Version two of the pre test self assessment survey will

contain one of the final video's, randomized to rotate between the male and female actor for each respondent. Respondents will be randomly assigned to version one or version two of the pre-test self assessment survey, resulting in half of the respondents being given a computer based text administration of the assessment survey, and half of the respondents being given a pre-recorded video administration of the pre test self assessment survey. All of the respondents who received version 2 with the pre recorded video will receive either a male or a female video assessment administrator. Both of these randomization procedures will be conducted utilizing the built in features of the MTurk software, with the intention of producing a randomized, balanced sample between version one and version two.

The BAROCO short syllogism test will be used to, and has been found to contain construct validity as an indication of general ability within a sample of people ranging between 13 and 68 located in Tokyo Japan (Shikishima et al., 2011a). Additionally, the BAROCO short has been found capable of serving as a proxy for intelligence testing (Shikishima et al., 2011a). The advantage of utilizing the BAROCO short as a means to have respondents demonstrate a task is that the answers to the BAROCO short cannot be easily found, improving the reliability and validity of the results. (Shikishima et al., 2011a). The BAROCO short can also be easily normed for different cultures or regions simply by replacing the common names written in the original version of the BAROCO short, to names common to the region of the world in which the respondent is participating (Shikishima et al., 2011a). The normalization based on one of the three geographic regions will occur prior to respondents being solicited, and will be a function

of the software to provide the correct version to respondents based on their preferred language.

Procedures

Respondents will first need to accept participation in the survey, titled “Discover your IQ in five short questions!”. Once they have selected the survey, the respondent will be given a standard informed consent, and compensation agreement prompt via text provided on the screen, and corresponding selection buttons. If the respondent provides informed consent, and elects to participate as a respondent, they will be shown one of the two versions of the pretest self-assessment survey. Respondents in the version one (text only) will be given the instructions on how to complete the assessment, and will be shown the statement “all answers are completely confidential, your participation is anonymous, and your scores will not be seen by anyone able to link a particular score to a particular person”. Respondents in version two will view instructions given by the pre-recorded video. The instructions will be verbatim of version one, but will be delivered by the pre-selected video assessment administrator. The video will then state: “all answers are completely confidential between you and I, your participation is anonymous and only I, the testing administrator, will see your score, no one else will be able to connect a particular score to a particular person, and I will keep your scores completely confidential.” Both version one and version two will then ask (via text in version one, via pre-recorded video in version two) the demographic questions, and the prediction strength question.

Here, anonymity is used as a means to moderate the extrinsic motivation positing that more extrinsic motivation will be present when a person is visually observed stating

that they will know your score. The risk with this use of anonymity as a means to moderate extrinsic motivation is that respondents will opt out of the study, rather than continuing to completion, if they feel that they are not doing well or will not do well on the intelligence test, possibly to avoid embarrassment. This will be mitigating by stating both in the pre-test self assessment survey, as well as in the informed consent document, that all scores will be kept confidential with the researchers, and will not be associated with a respondent by name.

Following the first prediction strength question, the BAROCO short will be displayed in full, with one question appearing on the screen at a time, immediately preceded by the pre-test self assessment questions. The preceding question will be asked prior to each of the five syllogism questions. Each of the five questions on the BAROCO short will be presented in the same order for each respondent. The prediction question asked prior to each of the five BAROCO short questions is:

"I think I will answer ___ out of____ questions right on this test"

a. 1 out of 5
b. 2 out of 5
c. 3 out of 5
d. 4 out of 5
e. 5 out of 5

The order of the answers available to this prediction question will be randomized to control for ordering effects. The question will be asked via text in Version 1, and via pre-recorded video in version 2. The BAROCO short will be presented via text in both version 1 and version 2. With the five BAROCO short questions, and the preceding

prediction question, this will total 10 questions during the assessment. After completing the 10 questions, the respondent will be given post test instructions that provide the actual answers to the syllogism test, contact information for the researcher, and a thank you message all via text in version 1, and via video message in version 2.

The research method proposed utilizes qualitative analysis, coded by quantizing accuracy of prediction as a means to represent overconfidence through the creation of a prediction index. Here, it is posited that overconfidence can be mediated during self assessments. In order to present the opportunity for overconfidence to be displayed, we present a task to be completed, the BAROCO Short Syllogism test, as well as a self-assessment prior to each question of the test. The self assessment asks how well one thinks they will do on the syllogism test by predicting their score. By first quantizing each of the prediction questions, and then comparing the predicted scores on the test to the actual score, a level of accuracy is created. By coding this level of accuracy, quantitative analysis are possible, such as the ANOVA

The accuracy of prediction score (*ap)* will be coded through a comparison between each respondents self-reported prediction of their score on the BAROCO short syllogism test (self assessment score, *sa)* and their actual score (actual score, *s)* on the BAROCO short test administered as part of this study. The BAROCO short syllogism test is made up of five questions (Shikishima et al., 2011a), and each question will be preceded by a single self assessment question asking the respondent to indicate how many questions out of the five question BAROCO short syllogism test they think they will get right. Each of the five self assessment questions regarding how many questions they will answer correctly are summed and an average score (*sa)* will be taken. This

self-assessment score (*sa)* will be stated as a fraction, ranging between 0 out of 5 to 5 out of 5. For example if a respondent predicts on each of the five prediction questions that they will score a 5 out of 5, there average self assessment score is 5 out of 5, or (*sa=1)*. In order to produce the accuracy of prediction score (*ap)*, coding will compare the self-assessment score (*sa)* to the actual score (*as)* of the respondent on the BAROCO short, by converting all scores to a decimal. The difference between these two scores will represent the respondents accuracy of prediction (*ap = as-sa)*, with 0 representing perfect accuracy of prediction, -1 representing predicting better outcomes than the actual result, and 1 representing predicting worse outcomes than the actual result. Here, scores closest to -1 would represent overconfidence in the respondent.

For example, if respondent number one average self assessment score is that they will answer 3 out of 5 of the BAROCO short questions correctly, the would receive a self assessment score of (*sa* = .6). If respondent number one actually answered 2 out of 5 questions correctly on the BAROCO short, they would receive an actual score of (*as* = .4*)*. In order to determine respondent number ones accuracy of prediction, the difference between the actual score and the self assessment score must be found (*ap* = .6-.4*), or (ap* = .2*)*. In this example, respondent number one's accuracy of prediction score is equal to .2.

In order to control for any ordering or learning affects associated with the prediction questions that immediately proceed each of the 5 BAROCO Short questions, an additional prediction strength question is given during the pre-test self assessment survey. This prediction strength question produces a prediction strength score (*ps)*, and is based on the response given to the initial question regarding the overall performance on

the upcoming assessment. The choices for selection were given as a Reichert scale with five choices, very likely to get all questions right coded as a one, and not likely at all to get all questions right coded as a 0. This general prediction strength question, it the resultant interval data will then be used to conduct a second Pearson bivariate correlation with the actual Score on the BAROCO short.

Data Analysis

Hypothesis number one states that respondents with elevated scores on the BAROCO short syllogism test will have an accuracy of prediction greater than the accuracy of prediction of respondents with lower scores on the same test. The accuracy of prediction score can range from -1 to 1. In order to examine the correlation between the accuracy of prediction, and the score on the BAROCO short, these two interval scores will be compared using a Pearson bivariate correlation analysis within SPSS. Additionally, the strength of prediction score will also be compared to the actual score on the BAROCO short using the same procedure. Further analysis will be possible utilizing SPSS to assess the correlation between the variables accuracy of prediction, strength of prediction and BAROCO short score with the following variables Accuracy of Prediction Score, Strength of Prediction Score, BAROCO Short Score, Level of Education, and Ethnicity. Utilizing SPSS will allow data entry that results in the collection and comparison of mean scores across the sample on each of the relevant variables. It will also allow for the use of a person bivariate correlation to replicate the findings of (Dunning et al., 2012) that found that intelligence test score correlates to accuracy of prediction. Additionally, an independent samples T test to compare version one and version two will help to identify any differences in the survey administration method.

Here the average strength of prediction score will be found for each of the two groups, and the independent samples T Test utilized within SPSS to compare the two groups to identify a significant difference.

This methodology is limited in the use of pre recorded video in order to simulate a live test administrator, and results that find an impact of the pre-recorded video may not be generalized to in person assessment administration. However, if pre-recorded videos designed with the intention of recreating in person interviews is found to impact the level of overconfidence displayed during self assessments, the use of pre-recorded videos may be preferable in the administration of self assessments, were overconfidence is not desired. Additionally, by demonstrating an impact of simulated in person interviews on self-assessment outcomes, the foundation for future research into similar confounds may be warranted.

Ethical Considerations

The proposed study seeks to explore the ability of an interview administered self assessment to mediate the display of overconfidence prior to an intelligence test. The population for this study is broad, and can include anyone ages 18-68 that has the opportunity, access, or potential to participate in a computer based research study given via the internet, and located in the West, Middle East, or East. The requirement to have access, as well as the general ability to read and operate a computer, in addition to the need for access to internet will likely limit my study to respondents with a average socioeconomic status in countries where internet access and computer literacy is common across the population. Although similar studies utilizing the BAROCO Short syllogism test, as used within my study, have utilized respondents under the age of 18, I have

chosen to only solicit respondents above the age of 18 to streamline the consent process, in hopes that this will increase my final sample size. Additionally, because my study is planned to be carried out in multiple languages, the available respondents will be limited to individuals who are literate in the pre-selected languages in which the assessment is written. These languages will be selected based on the number of people that speak each language in each of the three geographic global regions. The reason I plan to carry out this study across three distinct global regions is to control for the impacts of culture on the display of overconfidence, as found in Dunning (2018). Here, differences in the overconfidence displayed during self assessments was impacted based on the cultural background of respondents, with individuals who identified themselves as relating with eastern culture, demonstrating different levels of overconfidence when compared to individuals from western cultures (Dunning, 2018). The ethical considerations to be given in carrying out this research is to ensure that findings similar to Dunning (2018) that point to differences that appear to be impacted by cultural, gender, age, or other demographic characteristics are delivered in and unbiased and fair way, ensuring that statements reflecting personal opinions, beliefs regarding differences, or other information that could cause a negative impact to a respondent, are omitted and avoided. Additionally, although the BAROCO Short has been compared to intelligence testing, with findings to support the use of this short assessment in assessing some forms of intelligence (Shikishima, 2011a), respondents must be informed of the limitations of the assessment, and the generalizability of the results.

Although the study proposes replication among multiple cultures, the intention is not to identify cultural differences, but instead, control for those differences by demonstrating a

homogeneous impact of survey administration method, across all cultures, thereby expanding the generalization of the results to a larger populations. Care will be taken in displaying the results regarding any findings related to culture that are not relevant to the point of the study, to ensure to ethical conflicts arise.

Additionally, respondent's information will be kept confidential, and the respondents will be informed of this confidentiality at the onset of the assessment. Data gathering, analysis, etc. will be conducted in a way conducive with maintaining each respondents confidentiality through the use respondent numbers instead of names. No personally identifiable question responses will be presented during the assessment in order to ensure that respondents do not self disclose personally identifiable information within the data collected. These measures will help to ensure that the results on the assessment, as well as the level of overconfidence are not directly linked to a particular person by name.

Conclusion

Overconfidence has been demonstrated to be a viable topic to explore within survey administration. The wealth of research existing on this topic make it desirable as a method of measuring the prevalence of overconfidence in relationship to the administration of the BAROCO short, however learning more about overconfidence and how it can be moderated is also a desirable outcome. The outcomes of this proposed research stand to build on the research presented herein, identify other key research existing within these elements of self assessment and survey administration, and develop sound method to investigate the correlative nature of the variables accuracy of prediction, pre-test survey administration method, and *g*.

This topic of research endeavors to explore more deeply the flawed self-assessment, in order to gain a better understanding of the phenomena described as overconfidence, as well as self-assessments in general. Given the consistency with which overconfidence is displayed in self-assessments, are there confounding variables within the methodological approach utilized to establish this phenomenon? This research aspires to provide some insight into this question. As demonstrated here, when reviewing the research conducted that consistently identifies overconfidence in self-assessments, one variable of interest is the way in which obtaining the incidence of overconfidence is obtained. It is worth noting that in assessing overconfidence in aptitude testing, often times in past research, overconfidence is evaluated with a methodological approach that involves the use of a pre-test survey administered either by pencil and paper, or more recently, by computer generated text survey (Dunning et al., 2018) Research conducted in survey administration method has identified substantive differences between surveys that are administered via computer compared to those that are administered in a face to face condition (Cooper and Rowe, 1996). If the pre-test self assessment survey given prior to a measure of aptitude (BAROCO short) is administered via text on a computer screen, as opposed to being delivered by a video recording that places the respondent face to face with the testing administrator, it seems plausible that there could be a difference in the level of overconfidence measured. If there is a difference in the level of overconfidence measured, based on the method of pre-test self-assessment survey administration, then additional data regarding the complexities of self-assessments may be available as well as information regarding the effectiveness of specific measures implemented to dissuade overconfidence where it is not desirable, such as utilizing video based instructions,

warnings, or organizational training that emphasize, feature, or attempt to replicate face to face interactions with possible emphasis on extrinsic motivation. Information regarding the moderation of phenomena like overconfidence through the use of methodological measures such as video dialogue could offer valuable insight into the world of survey administration with in psychological science, as well as within organizational assessments of employees and leadership. If the prediction that survey administration method will moderate overconfidence is incorrect, this research still offers a valuable method to measure overconfidence, and could be utilized by psychological research as well as organizations seeking to determine a baseline level of the expression of overconfidence in individuals.

Organizations can experience negative impacts as a result of flawed, or inaccurate self-assessments made by their employees, or leadership (Dunning, et al., 2004). The proposed research endeavors to demonstrate first the presence of overconfidence in self-assessments and then the ability to modify that overconfidence through moderating the administration method of the pre-test self-assessment survey. This research into overconfidence serves to add to the body of research regarding overconfidence, as well as provide a means to reduce overconfidence in future self assessments. The reduction of overconfidence displayed during self-assessments as part of this research not only stands to benefit the reduction of overconfidence in future self-assessments, but also highlights the role of survey administration method in self-assessments.

References:

Couper, M. P., & Rowe, B. (1996). Evaluation of a computer-assisted self-interview component in a computer-assisted personal interview survey. *Public Opinion Quarterly*, *60*(1), 89–105. https://doi-org.ezproxy.snhu.edu/10.1086/29774

Bureau of Labor Statistics (BLS), 2020. Economic News Release. *Table B-3. Average hourly and weekly earnings of all employees on private nonfarm payrolls by industry sector, seasonally adjusted.* https://www.bls.gov/news.release/empsit.t19.htm

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106. https://doi-org.ezproxy.snhu.edu/10.1111/j.1529-1006.2004.00018.x

Dunning, D., Heath, C., & Suls, J. M. (2018). Reflections on self-reflection: Contemplating flawed self-judgments in the clinic, classroom, and office cubicle. *Perspectives on Psychological Science*, *13*(2), 185–189. https://doi-org.ezproxy.snhu.edu/10.1177/1745691616688975

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, *138*(2), 296–321. https://doi-org.ezproxy.snhu.edu/10.1037/a0026556

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, *106*(4), 766–794. https://doi-org.ezproxy.snhu.edu/10.1037/0033-295X.106.4.766

Klag, S. M., Creed, P., & O'Callaghan, F. (2010). Early motivation, well-being, and

treatment engagement of chronic substance users undergoing treatment in a therapeutic community setting. Substance Use and Misuse, 45(7–8), 1112–1130. doi:10.3109/ 10826080903499562.

C, J. I., & Wright, J. C. (2011). Measurement of self-enhancement (and self-protection). In M. D.Alicke & C.Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 472–494). New York, NY: Guilford Press.

Logg, J. M., Haran, U., & Moore, D. A. (2018). Is Overconfidence a Motivated Bias? Experimental. *JOURNAL OF EXPERIMENTAL PSYCHOLOGY-GENERAL*, *147*(10), 1445–1465. https://doi-org.ezproxy.snhu.edu/10.1037/xge0000500

Musto, M. (2019). Brilliant or bad: The gendered social construction of exceptionalism in early adolescence. *American Sociological Review*, *84*(3), 369–393. https://doi-org.ezproxy.snhu.edu/10.1177/0003122419837567

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi-org.ezproxy.snhu.edu/10.1037/0003-066X.55.1.68

Shikishima, C., Yamagata, S., Hiraishi, K., Sugimoto, Y., Murayama, K., & Ando, J. (2011b). A simple syllogism-solving test: Empirical findings and implications for g research. Intelligence, 39(2–3), 89–99. https://doi-org.ezproxy.snhu.edu/10.1016/j.intell.2011.01.002

Shikishima, C., Yamagata, S., Hiraishi, K., Sugimoto, Y., Murayama, K., & Ando, J.

(2011a). BAROCO Short. PsycTESTS. https://doi-org.ezproxy.snhu.edu/Full text; 999915934_full_001.pdf

Sripada, R. K., Bowersox, N. W., Ganoczy, D., Valenstein, M., & Pfeiffer, P. N. (2016). Self-determination theory and outpatient follow-up after psychiatric hospitalization. *Community Mental Health Journal, 52*(6), 662–666. https://doi-org.ezproxy.snhu.edu/10.1007/s10597-015-9929-6

Trochim, T. M., (2020). Research Methods Knowledge Base. *Conjoint.ly.* *https://conjointly.com/kb/#using-the-kb-in-a-course*

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi-org.ezproxy.snhu.edu/10.1177/1745691613518075